Experimentation and learning: The design of policy experiments and their learning effects, a conceptual framework and application to a case study from the Netherlands.

McFadgen, B.K.

2014

IVM Institute for Environmental Studies

Experimentation and learning

The design of policy experiments and their learning effects, a conceptual framework and application to a case study from the Netherlands.

Report Deliverable 2.1.3

B. McFadgen

Report R-14/41
21 November 2014

This report is released by: Prof. Dave Huitema

This report was commissioned by: Knowledge for Climate
It was internally reviewed by: Knowledge for Climate

IVM
Institute for Environmental Studies
VU University Amsterdam
De Boelelaan 1087
1081 HV AMSTERDAM
The Netherlands
T +31-20-598 9555
F +31-20-598 9553
E info.ivm@vu.nl

Knowledge for Climate
Programmabureau
Daltonlaan 400
3584 BK UTRECHT
The Netherlands
T +31 317 48 6540
F +31 6 2120 2447
E info@kennisvoorklimaat.nl

Contents

Abstract

Adaptive management as defined by Lee (1999) places learning at the heart of environmental governance and presents experimentation as a leading enabler of learning. This article calls for attention to so-called policy experiments, and posits a definition of a policy experiment that best suits the needs of adaptive governance. It is argued that an experiment is essentially a policy tool that simultaneously enables and tests innovation in a temporary space. Although learning is described as the key objective in adaptive governance, and experimentation has often been implied in suggestions to enhance such learning, little is known about the relationship between experimentation and learning at the conceptual level. Therefore we develop a conceptual framework composed with an institutional perspective to help address this situation. The framework suggests that initiators of experiments need to take various interrelated design decisions when setting up an experiment and we delineate three ideal types of experiments: the technocratic, boundary, and advocacy experiment.

These types are expected to produce different levels of learning effects, both within the experiment and the wider policy network. We tentatively suggest that technocratic experiments can be expected to produce high levels of cognitive learning, no normative learning and low levels of relational learning; and that the results from the experiment will be considered credible. Boundary experiments on the other hand create high levels of normative and relational learning, but lower levels of cognitive learning. They will score high on legitimacy and salience, but poorer on credibility.

Finally, advocacy experiments lead to some cognitive and relational learning, but low levels of normative learning. Their findings will suffer in terms of credibility and legitimacy, but the salience might, under certain circumstances, be strong. To assess the usefulness of our framework and to provide for an initial test of our theoretical propositions, the framework is applied to a case study of an experiment conducted in the Netherlands. Application of our framework there brings some surprising results, with potential implications for the academic debate on experiments and the science-policy interface more broadly. We conclude by sketching some limitations of our research and by suggesting further research priorities in the field.

1 Introduction

Until recently, climate governance has relied on developing measures that attempt to mitigate the effects of a changing global climate. Adaptation, however, is now gaining traction as a vital part of the preparation, and a greater understanding is needed about how to govern adaptively (Termeer et al 2011). Adaptive governance is a field that offers a potentially useful perspective, in particular the notions of experimentation and learning (Folke et al 2005, Cooney and Lang 2007, Huitema et al 2009).

The planet's ecosystems are characterised by non-reducibility, variability, and a collective quality (Dryzek 1987), and the uncertainty inherent in these dynamic conditions may be addressed in part by the reversibility and exploratory nature of public policy experimentation (Millo and Lezaun 2006). In many social-ecological systems governments and their policies play an important role in the prevention and resolution of climate change related issues, and policy experiments are specific temporary institutional arrangements that operate in the development phase of the policy cycle. An advantage of experiments is that their temporary nature allows policy makers to learn about the effects of an intervention without fully committing to a course of action, thereby reducing uncertainty while maintaining flexibility.

The notion of experimentation was appealing in the decades after the Second World War, because that period was optimistic about the capacity of governments to provide for collective goods, and much was expected from science as guidance for policy making (Merton 1973, Campbell 1998). However, when the social engineering role of the state and the positive role of science in policy making were confronted (Dryzek 1993), the notion of experimenting lost support too. In the policy sciences field, experiments were equated with a positivistic, non-reflexive approach to policy making, with questionable validity and serious ethical consequences (see especially Fischer 1995, Vedung 1998, Martin and Sanderson 1999, Sanderson 2002, Greenberg *et al.* 2003). Although there are areas of public policy (social policy, education, health) where experiments have continued to hold appeal over the years as the gold standard of evidence-based policy making (Haynes *et al.* 2012), it is interesting to see experiments emerging in the discussion on environmental policy and governance. Experiments unite the adaptive management and adaptive co-management literatures (see e.g. Lee 1993, Folke *et al.* 2005, Armitage *et al.* 2008, Huitema *et al.* 2009) as an approach that tests hypotheses about alternative management interventions, and they are increasingly seen as an important ingredient of a learning approach in innovations and transitions management (Huitema and Meijerink 2009, Berkhout *et al.* 2010, Voss and Bornemann 2011). While, in the adaptive management literature, experiments are equated with the provision of data on the effectiveness of interventions, this literature also suggests that experiments play a role in the work of 'shadow networks'- informal, collaborative networks- that pave the way for transitions in government policy (Gunderson, 1999).

This endorsement of experiments, not only in academic circles but also in practice, makes research on the actual effects of experiments increasingly necessary. This paper is one contribution and it explores the relationship between policy experiment design on the one hand and learning on the other, and proposes a framework that can be used to systematically analyse their interaction. Throughout the adaptive governance literature, experiments and learning are consistently linked; for example, Armitage *et al.* (2008) describe experiments as a learning mechanism, which can produce experiential and reflexive learning-by-doing processes. Experiments are expected to

quicken the pace of learning; however there are technical, resource, and ethical risks to bear with active experimentation. Lee (1999) and Walters and Holling (1990) also discuss experimentation as a form of learning-by-doing where theoretical propositions are tested in the field. However, building a framework that unifies the two concepts is novel, and even experiments themselves have seldom been analysed in a rigorous and systematic way (some exceptions are Greenberg *et al.* 2003, Vreugdenhil *et al.* 2010, Farrelly and Brown 2011, Castán Broto and Bulkeley 2013). Learning is understood here as the normatively desirable outcome of experimentation, and is therefore expected to achieve a lot, including resolving conflict and enabling decision making in uncertain situations (Diduck 2010). Despite this, little is written about prescriptions that bring these effects about, particularly in the writings on experiments. One source of potential understanding here is the social learning literature, where institutional factors such as information dissemination, diverse participation, and power diffusion are identified as enablers of learning (Mostert *et al.* 2007, Muro and Jeffrey 2012). These factors can be translated into features of experimental design and grouped in specific configurations, which is what is attempted below.

This article is structured as follows: first, the framework is laid out, introducing concept definitions and, based on the literature, the variables used for systematic comparison. It is contended that the notion of policy experiment is underdeveloped, and could use a dose of conceptual clarity. Second, institutional design choices are grouped by distinguishing 'ideal type' experiments, and a proposition regarding how these types relate to learning is outlined. Third, the framework is tested using a case study example of a policy experiment in the Netherlands. How the results meet expectations and what they mean for the validity of the framework is discussed. Fourth the limitations to this research are considered. The conclusion remarks on the case study as a policy experiment with ideas for future research.

2 Analytical framework

2.1 Experimenting with public policy

Despite being relevant to a variety of academic fields, public policy experiments are difficult to readily identify. Both conceptual and empirical analyses of experiments are rather limited, although currently a small body of literature is experiencing growth as experiments emerge as a tool to meet humanity's needs of addressing complex policy issues, such as water and climate change mitigation (Farrelly and Brown, 2011; Castán Broto and Bulkeley 2013). In their analysis of adaptive co-management prescriptions, Huitema *et al.* (2009) identify two interpretations of experimentation: an approach to management and as a research methodology. Adaptive management is therefore understood as both an experimental process (similar to Sabel and Zeitlin's 'experimentalist governance'- Sabel and Zeitlin, 2011) and a method to produce reliable and valid information for policy making by testing management interventions in practice (see Lee 1999, Cook *et al.* 2004, Armitage *et al.* 2008). This second, evaluative function also defines the experiment as understood in political science, where project interventions with a classic experimental design (i.e. with randomisation and control groups) are expected to provide reliable evidence of "what really works", and thereby lead to evidence-based policy making (Rondinelli 1993, Campbell 1998, Millo and Lezaun 2006, Stoker 2010, Haynes *et al.* 2012).

A third interpretation is possible: that of acting outside of established rules. This way of looking at experiments recognizes novelty, and advocates for the release of incumbent laws and processes to produce an experimental space where new innovations can emerge. For example, an experimental 'niche' provides a protected space to innovate, bring actors together, and build the means to upscale a project into the mainstream (Berkhout *et al.* 2010). Climate change experiments, as conceptualised by Castán Broto and Bulkeley (2013), are interventions that innovate and focus on learning, and are identified by the fact they sit beyond the processes of official policy making.

For adaptive governance, an experiment is expected to generate alternative solutions to emerging problems (Olsson *et al.* 2006), so testing new ideas 'outside of the ordinary' is vital. However, it is argued here that there is a certain level of commitment needed if initiators want to experiment; they must be open to failure. To enable the possibility of failure, an experimenter must establish some control; at the very least a monitoring programme that allows the evaluation of the proposal's effects. In this way, experiments are distinguishable from other innovative projects. Drawing together the notions described above, an experiment is thus understood as a temporary, limited space of observation and control that provides a degree of systematic assessment of innovative policy ideas. It needs to both enable and test innovative concepts. Reflecting on an experiment's relationship to policy, Millo and Lezaun (2006) identify different modalities of experiments; including pilot programs, regulatory impact assessments, and field trials, which can be applied as policy approaches, mechanisms, instruments, and concepts. Experimenting with policy may be direct, like adopting a completely new policy instrument; or indirect- like testing new concepts or management interventions where results will have an impact on deciding the intended course of a policy.

These applications should be captured in a broad understanding of a policy experiment, and bringing both the experimental and the policy strands together, this

article defines policy experimentation as “a temporary, controlled field-trial of a policy-relevant innovation that produces evidence for subsequent policy decisions”.

Recently in this journal, Munaretto and Huitema (2012) assessed how experiments impact learning in an analysis of Venice lagoon. They concluded that experiments have the potential to produce both factual and reflexive learning because of the repeated interactions of actors, which enables the actors to deal with uncertainty and change (Munaretto and Huitema 2012). The framework developed here takes this observation a step further by attempting to assess whether a variety of factors produce learning, and its structure is described in the following section.

2.2 Experiment ideal types

In order to understand how the design of an experiment influences its learning effects, we imagine that the institutional rules of an experiment are essentially choices to be made when designing the experiment. Institutional rules are presented in the rule typology developed by Ostrom (2005) in the Institutional Analysis and Development Framework, and include the boundary, position, information, choice, aggregation, payoff, and scope rules. Focusing on institutional rules allows us to compare experiments with quite different characteristics. Table 1 describes six of the rules as defined by Ostrom (2005) and links them to choices that may be made when designing experiments:

Table 1 Institutional rules and their relation to experimental design.

Rule type	Rule definition	Design choice
Boundary (entry or exit)	Delineates who participates by defining how participants enter the experiment and what makes them eligible.	Deciding on what makes a participant eligible to participate and how others are blocked from participating.
	Determines how open the access is to an experiment.	Deciding in what way an actor can gain access to the experiment- by invitation from the initiator or through making a request.
Position	Assigns participants with a position in the experiment. The minimum number of positions is one- that of participant in the experiment.	The number of participants allocated as initiator, designer, implementer, and evaluator. Decision whether to include a facilitator; whether there is more than one party financing the experiment.
Information	Information rules prescribe the subjects or types of information that can be communicated.	Whether the experiment considers non-expert information valid, such as experiences and local knowledge; or only expert information. Whether reflexive forms of information are generated or only instrumental forms.
	Rules that prescribe the channels of information flow and frequency of communication.	How open the communication is, how frequently information is distributed, and how it is distributed.
Choice	Rules that prescribe what actions a participant can or cannot perform at particular node points in the process.	How power is distributed among participants, how much influence they have over significant points in the experiment; being the design, implementation, and evaluation processes.
Aggregation	Prescribe how collective decisions are made by individuals. If two or more participants have decision rights, the collective decision can be made either by majority or by consensus.	How the experiment assigns control over decision making, the weighting of each participant's decision making power
Payoff	Manage the distribution of costs and benefits.	Where funding comes from.

Ostrom (2005) warns that indefinite institutional configurations can result from various possible settings of the rules, so for the purposes of this investigation, examples of experiments are categorised using ideal types (Weber 1968). The ideal types are developed as a heuristic device to anchor the rule configurations into three main groups, thereby structuring the empirical investigation. These three types are theoretically shaped by diametric approaches to policy development and analysis- the technocratic and interpretive approaches (Fischer 2007), as well as a model of the science-policy interface that classifies different roles of science in policy making (Pielke Jr. 2007). These ideal types underscore different ways the rules can be set, and are hyper-rational courses of action against which real-world examples can be

approximated to varying degrees (Weber 1968; Dryzek 1987). It is emphasised that as models, they cover the essential details of policy experiments, but are not all encompassing, as illustrated in the case study below. The remainder of this section describes the three types and Table 2 summarises the differences in rule settings between them.

Technocratic ideal type

The technocratic experiment follows the rational model of policy decision making where the technocratic elite generate expert knowledge for policy decisions (Fischer 2007). A discerning feature is the determined separation of facts and values, with the normative perspective decided at the outset by elected politicians or their policy advisors, and effectively tested in reality. The experiment then sidesteps conflicts in goals and values (Fischer 2007) although it is suggested that complete separation of fact and value is impossible to attain (Dryzek 1993, Fischer 1995). The role of a technocratic experiment in policy development resembles what is termed the 'pure scientist' or 'science arbiter' (Pielke Jr. 2007). In this arrangement, the experiment produces scientific information with little or no connection to the policy process until the end, when the results are presented to decision makers. This arrangement separates science from policy decision making and helps reinforce the view that science is impartial to politics, which upholds the scientific integrity of the evidence but may compromise its policy relevance.

Boundary ideal type

A boundary experiment represents a deliberative, inclusive design for producing policy relevant evidence. It is so named due to its role as a boundary object (Star and Grisemer 1989), engaging participants in a deliberative process and withdrawing their knowledge and normative perspective on the policy proposal being tested (Dryzek 1987). Ideally, this engagement will allow different interpretations of the policy problem to emerge that build into a common consensus on the most appropriate course of action. A range of state and non-state actors participate, and the knowledge of non-experts has value in the process, including anecdotal, traditional, and other contextual forms of shared information (Lindblom and Cohen 1979, Funtowicz and Ravetz 1990). There are no restrictions on what information is shared or who it is shared with, and all participants have authority to influence the problem definition, and design, monitoring, and evaluation choices. The role of the experiment resembles the 'honest broker of policy alternatives' (Pielke Jr. 2007), where it engages with the policy process and develops policy solutions in accordance with multiple value-perspectives.

Advocacy ideal type

An advocacy experiment generates evidence that supports pre-defined policy goals. Non-critical participants are selected who are involved in the policy issue and who already know each other. The initiator maintains control over the evidence produced so evidence that disputes the proposal is suppressed, and only knowledge that supports the proposal is shared. Policy goals are not up for debate since they were decided in advance by the initiator. In regards to its relationship with policy, the experiment resembles a '(stealth) issue advocate' (Pielke Jr. 2007) where it narrows the policy options to the preferred course of action. Campbell (1998) considers why initiators might choose to design an advocacy experiment and surmises that the effort of organising a project that confronts an established way of doing something risks 'institutional inertia' and thus any chance of failure is suppressed. Another explanation

is that an initiator believes they are right, and wants an opportunity to prove it, under the guise of "testing" something (see also Sanderson 2002).

Table 2 Policy experiment ideal types.

Rule type	Technocratic experiment	Boundary experiment	Advocacy experiment
Boundary	Participants are experts that contribute scientific or professional expertise. No procedures are established that will allow involvement of non-expert participants.	Procedures allow for scientific and professional experts, state actors, and non-state actors such as business and industry representatives and ordinary citizens, to participate in the experiment.	Rule settings allow for scientific and professional experts, state actors, and non-state actors such as business and industry representatives and ordinary citizens, to participate in the experiment, as long as they support the intervention and are known by the initiator.
	Entry into the experiment is by invitation from the initiator.	Entry is open to anyone who makes a valid request.	Entry into the experiment is by invitation from the initiator.
Position	No facilitator position is created.	A facilitator position is available.	No facilitator position is created.
Information	Produces expert-particularly scientific-information about the effectiveness of the intervention. Information is wholly instrumental.	Produces expert, and non-scientific (lay) information about the effectiveness of the intervention. Information is instrumental as well as reflexive (e.g. stakeholder perspectives on the suitability of the intervention for public policy).	Produces mainly non-expert (lay) information and restricted expert information that supports the pre-defined ends. Does not generate reflexive information.
	Channels of communication are open to all participants. Communication outside the experiment is minimal.	Channels of communication are open to all participants. Communication outside the experiment is regular.	Channels of communication are restricted among participants. Media reports are used to advocate the position.
	Information is distributed to participants using formal methods but not informal methods; e.g. no face-to-face workshops.	Information is distributed frequently using formal and informal methods; e.g. face-to-face workshops.	Information is not distributed to all participants, unpopular information is suppressed.
Choice	Initiator retains control over each decision node, with participants contributing their expertise when requested.	All participants have full decision making powers at each node point.	Initiator makes decisions at all nodes with minimal consultation with other participants.

Rule type	Technocratic experiment	Boundary experiment	Advocacy experiment
Aggregation	Majority decision making by a group of expert participants selected by the initiator.	Decisions are made by consensus.	Majority decision making by self-appointed group of participants.
Payoff	State initiator or research body pays all costs.	Experiment requests buy-in from all participants within reason.	Dominant party pays all costs.

In summary, the three ideal types of experiment each represent alternative methods of policy development with divergent configurations of participants, information, and power distributions. Experiments could be examined in other ways; for instance, by their differing impacts on the policy process (e.g. Greenberg *et al.* 2003), but by dissecting their institutional variables, learning effects can be systematically examined using pre-established factors that are said to influence learning. This is the strength of the framework, and it will be critically evaluated in the case study below. First, however, exactly what is meant by 'learning' is outlined.

2.3 Policy learning

Learning has an extensive and varied scholarship and in order to position the scope of our research, we draw on the dimensions of learning earmarked by Bennett and Howlett (1992): learning what, by whom, and to what effect. Regarding the first question, we understand that when someone learns, they are making sense of the world by acquiring new knowledge, or changing their perceptions at an individual level (Newig *et al.* 2010). Since the focus here is on the policy process and thus on policy learning, the following definition is adopted: "relatively enduring alterations of thought or behavioural intentions that result from experience and that are concerned with the attainment (or revision) of public policy" (Sabatier 1987). Following Haug *et al.* (2011) we distinguish three forms of policy learning: *cognitive learning* as the gaining of new knowledge and improved restructuring of existing knowledge; *normative learning* that captures an individual's deeper understanding of the policy process, and requires reflection on, and changes in, their perspective, goals, or priorities; and *relational learning*, which is defined as an increase in trust, an improved ability to cooperate, and a better understanding of the mindsets and frames of other participants. These categories are considered useful to empirically delineate between different learning evidence without conceptualising one as deeper, or more essential, than another. All are fundamental to environmental governance; for example, information is important to the policy process due to its advocacy and enlightenment functions (Grin and Loeber 2007, Sabatier 1978). High normative learning may indicate a synthesis of priorities between individuals about the policy issue and development of a common interest or goal within the group, leading to political consensus and collective action (Leach *et al.* 2013). Poteete, Janssen and Ostrom (2010) underline the importance of trust in governing social-ecological systems, and relational learning mirrors the goal of moral development (Webler *et al.* 1995) which enables participants to consider alternative perspectives and better cooperate with one another.

Regarding the question of "who learns?", the unit of analysis used here is learning at the individual level, because it is considered that only individuals can learn, although a culmination of their learning experiences can be taken as learning by the group as a

whole. Our research focuses on the learning effects in two different spaces. The first space consists of those engaged in the experiment itself- the participants of the experiment. Depending on the design of the experiment, this circle may contain just state actors or non-state actors as well. The second space involves state actors in the relevant policy environment, that is, the bureaucrats and decision makers engaged with the issue at hand.

As for the question "to what effect does learning take place?" there are different approaches in the literature on policy learning, with the dominant one equating learning with actual policy change (Bennett and Howlett 1992, Hall 1993, Busenberg 2001). However, it is recognised here that such change may occur without learning (i.e. with bargaining or negotiation), and learning can take place without policy change, such as being part of the development of shared understandings and mutual agreements (Leach *et al.* 2013). In connection with the questions on who learns and to what effect, there is a marked difference between learning amongst those involved in the experiment and learning amongst the wider circle of policy makers. The learning effects in the first group are expected to be relatively direct, and can thus be gauged immediately after an experiment has ended. The effects of experiments in policy circles, however, are likely to be indirect and protracted (see e.g. Weiss 1977). In order to gain insights in learning at the end of the experiment, the decision was made to use proxy indicators for learning that are known from the science-policy interface literature (particularly Cash *et al.* 2002). These indicators are the credibility, salience, and legitimacy of the experiment's findings on the relevant policy environment in the eyes of policy makers (Cash et al 2003). *Credibility* refers to the degree to which policy makers consider the findings of the experiment authoritative and believable, and to the degree in which they trust the outcomes. *Salience* refers to the relevance of the experiment findings at a certain moment in time. *Legitimacy* refers to the degree to which an information producing process was fair and whether it considered appropriate values, concerns, and perspectives of different actors (Cash *et al.* 2002.). These indicators are chosen here because they are well established in the literature, and because it is expected that fulfilling these criteria better will lead to higher learning effects in the longer run. Also, it has been suggested that it is difficult to maximize all three criteria (ibid.), and attempts to improve one criterion might actually lead to lower scores for the other criteria. This is attractive from the perspective of the current research as the various types of experiments might actually be geared to score well on different criteria and thus lead to learning effects amongst policy makers in diverging ways.

2.4 Relating experiment design to learning

The framework aims to assess how an experiment's design can influence cognitive, normative, and relational learning effects amongst the participants, as well as the perceived credibility, salience, and legitimacy of the experimental findings amongst policy makers. To build hypotheses on how experimental design and learning are related, use is made of the significant theoretical and empirical advances made in this journal that explain learning and suggest what sorts of variables, when brought together, will enable it (e.g. Mostert *et al.* 2007; Pahl-Wostl *et al.* 2007; Newig *et al.* 2010; Rodela 2011; Muro and Jeffrey 2012). For example, Mostert *et al.* (2007) outline a list of 71 factors in 8 themes that they found hinder or enable learning. The factors included independent facilitation, dissemination of information, joint planning and influence over the process, diverse but limited numbers of participants, common understanding, and frequent discussions. These factors have since been utilised as

independent variables in other learning studies (e.g. Muro and Jeffrey 2012), but for the purposes of this article, they are applied to the rules that describe an experimental design in particular. The main claim proposed by the framework is that since institutional rules can be configured so that they distinguish three experiment ideal types, and since, as explained above, learning is influenced by rule settings, the ideal types create different learning effects. Based on this and other work, the following hypotheses and supporting explanations are proffered:

Hypothesis 1: *A technocratic experiment produces high levels of cognitive learning, no normative learning, and limited relational learning within the circle of participants in the experiment. In the broader policy environment, the experiment's findings will be considered very credible, but they will not be considered salient or very legitimate.*

The reasoning underpinning this hypothesis is essentially that the emphasis of this experiment type is on building scientific expertise. The additional assumptions made are that the experiments will be initiated with the objectives already set, which means that an experiment is based on chosen points of debate within a paradigm, and that the theory tested in the experiment conforms to that paradigm. As the boundary rules preclude the entry of actors with different ideas, the information rules emphasize expertise that is fitting with that paradigm, as well as emphasise progress within the established paradigm, the experiment design precludes normative learning and produces learning that is mostly factual (cognitive). Due to the information rules allowing open communication and not suppressing information, there might be some relational learning in the process, but the form that this takes is more internal-scientific (who is best at solving the puzzle at hand, who can understand the system best) and not about creating higher levels of trust in a policy network. As for the perception of such experiments by the policy world, the funding for the experiment (pay-off rules) is likely to be from organizations with a purely scientific interest, that do not care greatly about policy relevant outcomes but more about scientific publications. The implication is that any salience of the outcomes is therefore probably a coincidental outcome. The closed character of the experiment makes the legitimacy of the results questionable as the research question, data gathering process, and report writing has not involved stakeholder groups or ordinary citizens and might not address arguments they consider important. A high level of credibility is expected here, in the very simple assumptions that scientific information is developed according to the highest standards of reliability, and that scientific information is perceived to be credible in general.

Hypothesis 2: *A boundary experiment produces medium levels of cognitive learning and high levels of normative and relational learning within the circle of participants in the experiment. In the broader policy environment, the experiment's findings will be considered moderately or very salient and very legitimate, but not very credible.*

This hypothesis is based on the open design of a boundary experiment, where the boundary rules are set so broadly they include all stakeholders that want to participate. This enables participant diversity, which is considered a trigger of learning in learning analyses (Schusler *et al.* 2003, Mostert *et al.* 2007, Muro and Jeffrey 2008, Gerlak and Heikkila 2011, Leach *et al.* 2013). Participants are exposed to a variety of ideas and understandings of the policy problem, which are shared amongst the group through open and transparent information rules. Through the authority rules, non-state actors influence the setting of the problem definition and experiment goals, which increases the probability of public support for the intervention and the

generation of shared norms. Since there is no dominant paradigm through which to view the problem, a kaleidoscope of perspectives shapes the experiment paradigm, creating possibilities for normative learning. The focus on capturing different knowledge types- e.g. non-expert lay knowledge about the system within which the experiment is embedded- enhances the breadth of understanding about the experiment's effects, but undermines understanding of in-depth complexities, thereby hampering cognitive learning. Participants have decision power through authority rules to influence the evaluation process, thereby capturing a wide variety of concerns and generating trust in the political process- relational learning- through participants feeling their needs are being met (Dryzek 1987, Webler *et al.* 1995, Mostert *et al.* 2007, Muro and Jeffrey 2012). Within the policy network, boundary experiments are perceived as very legitimate, because the focus on stakeholder interests and establishing ways to meet those interests ensures the findings meet the expectations of society; however, the drawback of this design is that non-expert knowledge undermines the technical quality of the experiment findings meaning credibility is not perceived as very high. If the policy network is 'in tune' with the current norms of society, then the findings will be considered very salient, but if the policy network is responding to cues other than societal norms (e.g. international, economic, political influences), the findings will be moderately salient.

Hypothesis 3: An advocacy experiment produces medium levels of cognitive learning, low normative learning, and low to medium levels of relational learning. In the broader policy environment, the experiment's findings will be considered salient (under some circumstances) but they will not be considered very credible or legitimate.

This hypothesis reflects the intention of the advocacy experiment to demonstrate a pre-defined policy solution as the most suitable course of action. Boundary rules allow entry to a potentially diverse set of participants, but access is limited to those chosen by the initiator, restricting eligibility to those who support the proposal. Diverse actors contribute knowledge (Hegger *et al.* 2012, Muro and Jeffrey 2012) so there is some cognitive learning, and persuasion tactics might trigger slight normative learning (Haug *et al.* 2011), but not the same levels that sharing a breadth of viewpoints might (Schusler *et al.* 2003). Both types of learning are inhibited by the lack of open and regular lines of communication (Muro and Jeffrey 2012). Due to the participants knowing each other and having similar views, there is potential for some relational learning, but no new actor networks emerge and the suppression of certain information and lack of authority for most participants inhibits trust building. The perception of an advocacy experiment is that its credibility and legitimacy are questionable. Credibility is undermined by the cherry-picking of information, which affects the reliability of what is produced. The initiator's attempt to advocate for a particular proposal blocks participation by critical actors and undermines their concerns, limiting fairness and the perceived legitimacy of the project. However, the salience of the findings may be perceived of as high when the experiment acts as a means for keeping an idea alive (Greenberg *et al.* 2003), and outcomes are presented when the time is right.

Now that the model has been explained by connecting learning outcomes to the ideal types, the following section tests it for robustness against a case study set in the Netherlands.

3 Case study application of the framework to an experiment in the Netherlands

This case study serves to illustrate how the above conceptual framework can be applied empirically. The experiment puts into practice a change in approach to water management- combining flood defence and nature conservation- and was conducted in the Dutch Delta's Oosterschelde (see Figure One), a former estuary in south-west Netherlands. The Oosterschelde is a designated national park and a Natura 2000 site of European importance. Recognition of national significance is bequeathed because the estuary sandbars, intertidal flats, and salt marshes provide foraging and resting places for migratory birds and other wildlife. The estuary is also the site of a large storm surge barrier- the "Oosterscheldekering". This flood protection barrier is part of the Delta Works, a vast engineering programme undertaken to protect the Netherlands from flood risk. The intertidal flats also minimise wave damage to the barrier by buffering it during storm surges. Unfortunately, the storm surge barrier has inhibited the tidal flow in and out of the estuary, which in turn disrupts the erosion and replacement of sand so the flats are slowly drowning (Climate Proof Areas Brochure 2010) putting the wildlife habitat at risk and potentially increasing damage to the flood barriers.

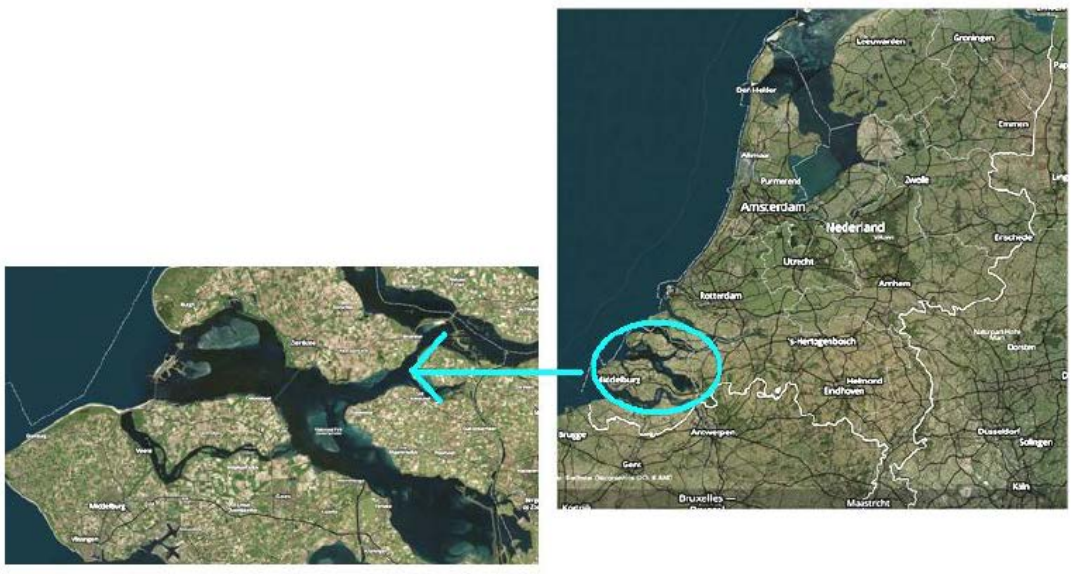

Figure 1 *Map of case study area (map sourced from © OpenStreetMap contributors).*

There is no precedent to managing these issues, so the enforcement arm of the Ministry of Infrastructure and Environment, Rijkswaterstaat (Department for Waterways and Public Works; referred to here as "RWS"), together with knowledge consortium Ecoshape, experimented with "soft approaches", whereby they tested innovative management interventions that can potentially slow down the rate of erosion; including a sand nourishment (experiment A) and the construction of sand stabilising oyster reefs (experiment B). These projects are a departure from the typical "hard" engineering measures used by RWS to maintain the dikes, and they hypothesised that the measures will reduce dike maintenance costs while restoring nature. The

intervention (experiment A and B together) stands out as an innovative experiment that will help determine whether or not RWS takes the soft solution approach, and manage dike safety and ecological restoration together as part of their water management planning in the future (RWS 2009), thus classifying it at as a management intervention specifically designed to inform policy (Steffen 2009; Huitema *et al.* 2009). The experiment's monitoring mechanisms were scientific-i.e. the morphological and ecological effects of sand nourishment and oyster beds, but there was also an objective to engage local stakeholders and build trust and legitimacy into the project. Permits were required because the experiment was conducted in a Natura 2000 protected area.

3.1 Methodology

To collect data, documents such as project reports, media articles, scientific articles, conservation reports, meeting minutes, email correspondence, and regulatory applications were analysed and six semi-structured interviews were conducted on the phone or in person. Data was also collected via an online survey. A pilot study to test the survey was conducted on professionals known by the researchers. The survey consisted of closed questions, but allowed respondents the opportunity to comment in an open section at the end. Questions were asked to gauge the institutional factors and learning effects. For this case study, 25 survey links were sent to the identified participants in the projects, and 20 responses were collected, giving an 80% response rate. Learning can be measured a number of ways, and a common method is to assess observed products of learning; for instance, policy changes, new projects, or new strategies (Bennett and Howlett 1992, Armitage *et al.* 2008, Gerlak and Heikkila 2011). An alternative method is to measure the process of learning, where new information or knowledge is acquired, processed, and transferred across individuals within a group (Gerlak and Heikkila 2011), and measured via each individual "reporting" their experiences through survey or interview. This research utilised the second method via a survey with experiment participants. Learning effects were measured with questions based on those found in the literature (e.g. Schusler *et al.* 2003, Muro and Jeffrey 2012, Leach *et al.* 2013), using Likert-scale survey questions, which asked respondents to rate their experience on a four point scale: 1= "not at all", 2= "slightly", 3= "a moderate change", 4= "a considerable change". The mean is provided below to assess the extent of learning effects from the experiment. The extent of learning in the policy network was assessed using indicators of the credibility, salience, and legitimacy, along with interview and survey data from policy actors and non-state participants.

3.2 Results and discussion

In order to establish whether the experiment fits any of the aforementioned ideal types, this section teases out the rule settings in accordance with the above framework. It also assesses whether the measured learning effects meet the expected hypothesis. Discussion points are then raised about the suitability of the framework to assess learning, along with observed limitations and suggestions for future research.

Table 3 Rule settings for the experiment.

Institutional rules	Results
Boundary rules	Access granted to scientific experts, state actors, and industry actors. State actors represent national and regional government. Participants invited to join.
Position rules	State was initiator and primary funder.
Information rules	Information produced was predominantly scientific information. Some lay information used in design phase of experiment B. Emphasis was on instrumental information and reflexive information extended to awareness of policy goals. Communication channels reported as open and information regularly received. Knowledge transmitted to general public through press releases. Participants had face-to-face contact through workshops.
Aggregation rules	State ultimately made decisions on project goals and design. The state took decisions on the execution of experiment A; the expert collaborative on experiment B. Evaluation and conclusions decided upon consensually by state and scientific experts.
Choice rules	The majority of participants reported holding advisory or decision making power roles at the monitoring and evaluation nodes. The least equitable power balance was at the design node.
Payoff rules	State paid most costs with contribution from research collaboration.

Our empirical analysis indicates that the experiment's rule settings most resemble the technocratic ideal type. The boundary rules are set heavily in favour of scientific experts, with minor representation of business and policy interests. However, on closer analysis, it emerged that civil society actors were invited into the process, but decided to take an observational stance (although a civil group is involved in the steering committee that received the results, and advised decision makers). They viewed the proposal as positive for both safety and nature, whereas industry actors perceived the intervention as potentially negative on their interests. Concerns were raised over impacts on recreational uses of the Oosterschelde if the management approach was formally adopted into policy but for the experimental phase civil groups did not feel the need to participate (Joop Stalenburg, telephone interview, April 2013). It emerged that participants were already very familiar with each other, as the Oosterschelde is a system supporting a balance of sensitive interests and the participants have all been working there for some time. The research institutes have a strong working relationship with the fishermen and they regularly pass knowledge between each other, so despite the project being innovative, the actor network was not. Examining information rules, the results indicate this experiment had an open and regular exchange between participants and open distribution to the public as a whole through the BwN's open access "wiki" site. However, on closer analysis the exchange of information was highest among scientific experts, and most open during the design of the experiment, which skews the categorisation towards the technocratic type, because information was not shared evenly across the whole circle of participants. The experiment investigated scientific hypotheses, so scientific information about the natural world was dominant; however, non-scientific lay information was used to design and implement experiment B, since the scientists had limited knowledge of oysters and useable substrate materials (Ysebaert interview, Yserke, December 2012). Both advocacy and boundary types value lay knowledge, although advocacy experiments limit it to that which supports the dominant interest. It is hard to evaluate in this instance because the fishermen's knowledge was practical and non-contentious. The experiment did not encourage the development of reflexive knowledge, although

some participants recalled discussion over policy objectives. Finally, regarding the authority rules, graph 1 below illustrates the distribution of authority at the three stages; the design, monitoring, and evaluation nodes. Degrees of authority were assigned as: not involved, received information, participant had advisory role, participant had decision power. We conclude from the graph that the evaluation node had more distributed power than the design or monitoring phases, and the design phase had the least shared power. Closer analysis shows that state actors were mainly responsible for the design phase, with few experts having any say at all (hinting at a technocratic or advocacy type), whereas scientific experts held decision power over the monitoring and evaluation stages.

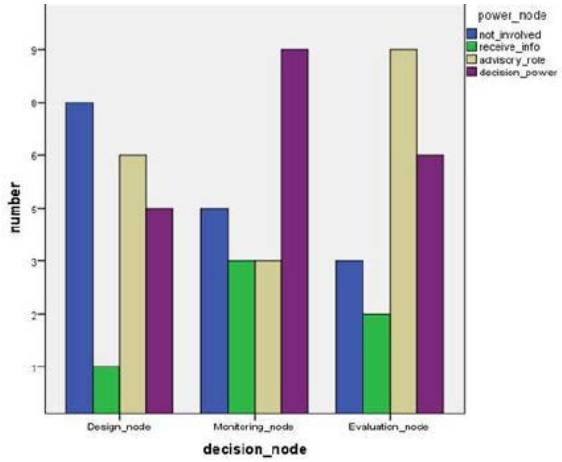

Figure 2 *Shows the extent that power was shared over decision nodes in the experiment. The columns show the numbers of participants with differing degrees of authority (power nodes) which illustrates the extent of their authority to influence the experiment at each of the design, monitoring, and evaluation stages.*

In conclusion, since the rules are set to include mostly scientific experts who are assigned authority at nodes of their expertise, and who regularly and openly sharing scientific and non-reflexive information, it is concluded that on the whole this experiment is primarily a technocratic ideal type. In regards to learning, the hypotheses state that a technocratic ideal type creates high cognitive learning effects, no normative learning, low relational learning and results that are considered credible by policy makers. The results are surprising (see Table 4 below) in that some normative and relational learning was reported; in particular the relational learning was unexpectedly high. The levels may reflect the fact that the experiment is not a perfect fit to the ideal type, as it had broader authority distribution than expected, and sharing of reflexive knowledge. The results also indicate how some rule settings have blanket influence over all types of learning. For example, the open and regular sharing of information increases both relational and cognitive learning, as participants would get to understand how others interpret the problem while increasing their knowledge of the experiment. We have not found examples of studies that isolate the variables that influence specific types of learning, and this could be an interesting direction for learning scholarship to take. Moreover, the institutional rules may not be extensive enough to explain the normative learning, which may be the result of forces beyond the institutional setting - the involvement of a charismatic leader for example (Gerlak and Heikkila 2011).

Table 4 Learning measurements from the survey, where participants responded to questions about the change in their understanding, views, and relationships.

Cognitive learning	3.3
Normative learning	2.5
Relational learning	3.0

(Out of 4: where >3.0 is high; 2.5- 3.0 is medium, and <2.5 is low).

Finally, an assessment is made of the quality of the experiment's results against the proxy learning indicators *credibility*, *salience*, and *legitimacy*. To understand the extent to which the results were credible, they were examined to see whether there were any critical questions raised from the policy side. The initiator reported no criticism or questioning of results, and suggested policy makers recognized the caliber of the knowledge institution contracted to perform the monitoring work (van Zanten, interview, March, 2014). In addition, larger sand supplementations are now planned for the Oosterschelde based on these results, which illustrates the confidence decision makers have in the results of this experiment (in line with findings from Greenberg *et al.* 2003 that policy makers view scientifically established results as definitive). To establish how salient the results were, we looked at whether there has been a change in how relevant the policy issues are after five years, and whether there has been a suggested alternative course of action. What we found is that Dutch water governance is still intent on pursuing solutions that solve nature and water safety policy problems concurrently, for instance as mentioned in the south-western Delta programme. By focusing on improving safety while restoring nature, the government can require buy-in from conservation groups, and only spend state money on safety solutions (van Zanten, interview, March, 2014). We also consider the collaboration with consortium Ecoshape as a strengthener of the experiment's policy relevance, as it transformed the experiment from a state-run safety and nature project to a highly visible and well publicized demonstration of a fashionable principle. So, there was no proposed alternative action, but rather an upgrading of the original project into something more visible. Finally, the legitimacy of the results can be understood as how equitable the process was, in particular, how the stakeholders' concerns were met. Here too, the experiment scores well, as it developed with industry stakeholders a reporting system that would alert participants to adverse effects caused by the sand supplement, with a ceiling effect arranged so operations could be halted if the effects proved too much. The extent of industry involvement beyond appeasement is less clear, as they were not involved in the final workshop, evaluation, or presentation of results. However, the Oosterschelde National Park had a representative sitting on the steering committee, and they felt satisfied with the extent stakeholders were involved in the process (Kramer, email correspondence, March, 2014).

On the whole the experiment is strong regarding the indicators for effectiveness. It is acknowledged that these results may be explained not only by the experimental design, but also by external factors; for example, the political awareness of climate change solutions for the Netherlands (sand supplementation and oyster beds were touted as climate change solutions in the EU programme "Climate Proof Areas"), and the strong support society gives to projects that improve safety. Further, as with relational learning, assessing legitimacy neglects to take into account how established the actor network is in the area prior to the experiment, and managing the Oosterschelde has long been an example of the Dutch "polder model" of consensus-based management.

It is concluded that despite not fitting perfectly, the framework captures a range of institutional factors that potentially enable learning. Further, the real world application shows both the depth of analysis achievable, and allows the framing of recommendations. However, an important 'warning' observation about using rules to define independent variables is the expectation that they are choices to be made, but although in some cases they are not. For example, with the boundary rules in this case, legal requirements (being situated in a Natura 2000 area) meant the initiators were obliged to invite all stakeholders into the process. The regulations would make it very difficult for the initiators to keep particular parties out. Conducting the experiment in a non-restricted area would have solved this, but then the state would not be compelled to find a solution in the first place. On the whole, the rules offer choices to designers, and although a charismatic leader would be desirable, it is not that easy to choose for. By restricting our analysis to institutional factors, our recommendations give designers some control.

3.3 Limitations

The exploratory research presented here has a number of limitations that also suggest directions for further research. The first limitation we wish to mention is conceptual in nature. Acting on the claim that the literature on experimentation is still somewhat weak or inconsistent, a systematic institutional approach was deliberately proposed. This approach functions well in anchoring the typology of experiments. However, an institutional lens is necessarily limiting, and future research should also pay attention to, or control for, other factors that might have relevance for explaining the learning effects. For example, control for what could be labeled 'agency factors' in the experiment circle: the quality of the leadership of experiment initiator; the demographics of participants, extent of motivation etc. (see for instance Gerlak and Heikkila, 2011; Leach *et al.* 2013). Regarding learning in the policy circle, legislation that may or may not curtail the space for experimentation has already been alluded to. But the relations in the policy system may also vary considerably – the parties involved (e.g. regulators and interest groups) may be either in an antagonistic or in a collaborative relationship. This might be quite important for the type of experiment that is feasible, but it will certainly also have implications for the perceptions of the outcomes of the experiment in policy circles.

In terms of the empirical approach used here, several limitations need mentioning. The first is that the analysis outlined here is based, at least partially, on self-reported insights into learning, measured after the experiment only. Self-reported learning scores are recognized as likely to be biased to some extent (Haug *et al.* 2011), and measurements of learning effects should preferably be based on pre- and post-experiment data, which were unfortunately not available for this study, as there was a need to focus on completed experiments. One future avenue of research would be to apply the framework to an experiment that is yet to start, and take learning measurements before and after the experiment (see for instance Huitema *et al.* 2010). The current application to just one case study is suitable for a test application, but there is a need for additional quantitative and qualitative work. We intend to follow up the current article with the construction of a large scale database that contains multiple experiments. On the basis of such a database, statistical analysis of the correlation between experiment design and learning effects can be established more firmly. And further qualitative research would be necessary – for instance in a select number of cases of experiments – to better assess what are the casual dynamics underlying the patterns of correlation that are found. Such research could also pay

attention to the connections between learning within the experiment on the one hand, and learning effects in the wider policy environment on the other; for example, do strong learning effects in the experiment lead to a positively perceived impact in the policy network? If not, why not?

4 Conclusions

Policy experimentation is a concept that is gaining importance in academic and policy discourse, due to its function as an institutional arrangement that promises to improve policy and management in the context of social-ecological uncertainty and feedback. Yet experimentation has quite specific meanings in different scientific fields, ranging from testing the effects of a policy innovation, to capturing an innovative and novel approach that sits outside the status quo. A useful definition should capture both the evaluative and novel aspects of the concept, and set the course for identification and analysis of experiments for future research, as well as inform policy discussions and governance strategies that want to apply a more adaptive approach to policy making. As an innovative framework that connects experiments and learning, it is argued that on the basis of the case study that the design features identified are relevant to different learning effects, and can contribute to future systematic assessments of learning within experiments, as long as the identified concerns (e.g. the lack of some choices, drawbacks of self-reported learning, focus on institutional factors) are acknowledged. It is concluded that the case study in question, a policy experiment assessing the impact of taking the “soft approach” by combining coastal defence and ecological restoration, adequately captured the temporary and flexible nature of the concept, and enable an understanding to be developed of how experiments can be used to both test a scientific approach, as well as gauge how the social world responds to the suggested policy change. It also highlighted the fact that policy approaches that are supported by the local population may not draw parties into the process despite open access, thus minimising the chances to learn from local actors.

Sketching a conceptual relationship between experimentation and learning is the first step in building a robust understanding of how experiments can be designed to enhance learning outcomes. The next step, testing, will validate the carefully constructed claims. This explorative research is considered timely because it unpacks a relationship between two concepts, which are often held up as an antidote to policy making by spurious certitude. As key prescriptions in adaptive management, experimentation and learning enable us to better cope with the uncertainties that arise in modern environmental governance.

5 Literature cited

Armitage, D., Marschke, M. & Plummer, R. (2008). Adaptive co-management and the paradox of learning. *Global Environmental Change, 18*, 86–98.

Bennett, C.J. & Howlett, M. (1992). The lessons of learning -reconciling theories of policy learning and policy change. *Policy Sciences, 25*(3), 275–294.

Berkhout, F., Verbong, G., Wieczorek, A.J., Raven, R., Lebel, L. & Bai, X. (2010). Sustainability experiments in Asia: Innovations shaping alternative development pathways? *Environmental Science & Policy, 13*(4), 261-271.

Busenberg, G.J. (2001). Learning in organizations and public policy. *Journal of Public Policy, 21,* 173–89.

Campbell, D.T. (1998). The Experimenting Society. In Dunn, W. (ed.), *The Experimenting Society: Essays in Honor of Donald T. Campbell. Policy Studies Review Annual volume 11* (p. 35). New Brunswick, New Jersey: Transaction Publishers.

Castán Broto, V. & Bulkeley, H. (2013). A survey of urban climate change experiments in 100 cities. *Global Environmental Change, 23*(1), 92-102.

Cash, D., Clark, W., Alcock, F., Dickson, N., Eckley, N. & Jäger, J. (2002). *Salience, credibility, legitimacy and boundaries: Linking research, assessment and decision-making*. John F. Kennedy School of Government Harvard University Cambridge, MA, Faculty Working Paper RWP02-046.

Cash, D.W., Clark, W.C., Alcock, F., Dickson, N.M., Eckley, N., Guston, D.H., Jäger, J. & Mitchell, R.B. (2003). Science and technology for sustainable development special feature: Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 8086-8091.

Climate Proof Areas brochure (2010). http://www.climateproofareas.com/project/sand-nourishment

Cook, W.M., Casagrande, D.G., Hope, D., Groffman, P.M. & Collins, S.L. (2004). Learning to roll with the punches: adaptive experimentation in human-dominated systems. *Frontiers in Ecology and the Environment, 2*(9), 467–474.

Cooney, R. & Lang, A. (2007). Taking uncertainty seriously: Adaptive governance and international trade. *The European Journal of International Law*, *18*(3).

Didick, A. (2010). The Learning Dimension of Adaptive Capacity: Untangling the Multi-level Connections. In Armitage, D. and R. Plummer (Eds.), *Adaptive Capacity and Environmental Governance (*p. 199). Berlin, Heidelberg, Germany: Springer-Verlag.

Dryzek, J. (1987). *Rational Ecology. Environment and political economy*. New York, USA: Basil Blackwell.

Dryzek, J. (1993). Policy analysis and planning: from science to argument. In Fischer, F. & Forester, J. (eds.), *The argumentative turn in policy analysis and planning (*pp. 213–233). Durham: Duke University Press.

Farrelly, M. & Brown, R. (2012). Rethinking urban water management: Experimentation as a way forward? *Global Environmental Change, 21*(2), 721–732.

Fischer, F. (1995). *Evaluating Public Policy*. Chicago, Illinois, USA: Nelson Hall.

Fischer, F. (2007). Deliberative policy analysis as practical reason: Integrating Empircal and Normative Arguments. In Fischer, F., Miller, G.J. & Sidney, M.S. (eds.), *Handbook of Public Policy Analysis: theory, politics, and methods* (pp.223-237). Florida, USA: Taylor & Francis Group.

Folke, C., Hahn, T., Olsson, P. & Norberg, J. (2005). Adaptive governance of social-ecological systems. *Annual Review of Environment and Resources, 30*(1), 441-473.

Funtowicz, S.O. & Ravetz, J.R. (1990). Uncertainty and quality in science for policy. Dordrecht, The Netherlands: Kluwer Academic.

Gerlak, A. & Heikkila, T. (2011). Building a theory of learning in collaboratives: Evidence from the Everglades restoration program. *Journal of Public Administration Research and Theory, 21*, 619-644.

Greenberg, D., Linksz, D. & Mandell, M. (2003). Social experimentation and public policy *making.* Washington, D.C., USA: The Urban Institute Press.

Grin, J. & Loeber, A. (2007). Theories of policy learning: Agency, structure, and change, In Fischer, F., Miller, G.J. & Sidney, M.S. (eds.), *Handbook of Public Policy Analysis: theory, politics, and methods* (pp.201-219). Florida, USA: Taylor & Francis Group.

Gunderson, L. (1999). Resilience, flexibility and adaptive management – antidotes for spurious certitude? *Conservation Ecology, 3*(1), 7.

Hall, P.A. (1993). Policy paradigms, social learning, and the state: the case of economic policymaking in Britain. *Comparative Politics, 25*(3), 275-296.

Haug, C., Huitema, D. & Wenzler, I. (2011). Learning through games? Evaluating the learning effect of a policy exercise on European climate policy. *Technological Forecasting and Social Change, 78*(6), 968-981.

Haynes, L., Goldacre, B. & Torgerson, D. (2012). *Test, learn, adapt: Developing public policy with randomised controlled trials*. London: Cabinet Office.

Hegger, D., Lamers, M., van Zeijl-Rozema, A. & Dieperink, C. (2012). Conceptualising joint knowledge production in regional climate change adaptation projects: success conditions and levers for action. *Environmental Science & Policy, 18*, 52-65.

Huitema, D. & Meijerink, S. (eds.) (2009). *Water policy entrepreneurs: a research companion to water transitions around the globe.* Cheltenham, UK: Edward Elgar Publishing.

Huitema, D., Mostert, E., Egas, W., Moellenkamp, S., Pahl-Wostl, C. & Yalcin, R. (2009). Adaptive water governance: assessing the institutional prescriptions of adaptive (co-) management from a governance perspective and defining a research agenda. *Ecology and Society, 14*(1), 26.

Huitema, D., Cornelisse, C. & Ottow, B. (2010). Is the jury still out? Toward greater insight in policy learning in participatory decision processes—the case of Dutch citizens' juries on water management in the Rhine Basin. *Ecology and Society, 15*(1), 16.

Leach, W.D., Weible, C.M, Vince, S.R., Siddiki, S.N. & Calanni, J.C. (2014). Fostering learning through collaboration: knowledge acquisition and belief change in marine aquaculture partnerships. *Journal of Public Administration Research and Theory, 24*(3), 591-622. doi:10.1093/jopart/mut011.

Lee, K.N. (1999). Appraising adaptive management. *Conservation Ecology, 3*, 3-16.

Lindblom, C.E. & Cohen, D.K (1979). *Usable knowledge: social science and social problem solving.* Yale University Press, USA.

Martin, S. & Sanderson, I. (1999). Evaluating public policy experiments: measuring outcomes, monitoring processes or managing pilots? *Evaluation, 5*(3), 245-258.

Merton, R.K. (1973). Sociology of Science. Chicago, USA: University of Chicago Press.

Millo, Y. & Lezaun, J. (2006). Regulatory experiments: genetically modified crops and financial derivatives on trial. *Science and Public Policy,* 33(3), 179-190.

Mostert, E., Pahl-Wostl, C., Rees, Y., Searle, B., Tàbara, D. & Tippett, J. (2007). Social learning in European river-basin management: barriers and fostering mechanisms from 10 river basins. *Ecology and Society, 12*(1), 19.

Munaretto, S. & Huitema, D. (2012). Adaptive comanagement in the Venice Lagoon? An analysis of current water and environmental management practices and prospects for change. *Ecology and Society, 17*(2), 19.

Muro, M. & Jeffrey, P. (2008). A critical review of the theory and application of social learning in participatory natural resource management processes. *Journal of Environmental Planning and Management, 51*, 325–44.

Muro, M. & Jeffrey, P. (2012). Time to talk? How the structure of dialog processes shapes stakeholder learning in participatory water resources management. *Ecology and Society, 17*(1), 3.

Newig, J., Günther, D. & Pahl-Wostl, C. (2010). Synapses in the network: learning in governance networks in the context of environmental management. *Ecology and Society, 15*(4), 24.

Olsson, P., Gunderson, L.H., Carpenter, S.R., Ryan, P., Lebel, L., Folke, C. & Holling, C.S. (2006). Shooting the rapids: navigating transitions to adaptive governance of social-ecological systems. *Ecology and Society, 11*(1), 18.

Ostrom, E. (2005). *Understanding Institutional Diversity*. New Haven, Connecticut, USA: Princeton University.

Pahl-Wostl, C. (2006). The importance of social learning in restoring the multi-functionality of rivers and floodplains. *Ecology and Society, 11*(1), 10.

Pelikán, P. & Wegner, G. (2003). *The evolutionary analysis of economic policy*. Northampton, MA, USA: Edward Elgar Publishing.

Peters, G.B. **(**1998). *The Experimenting Society and policy design*, p125. In **Dunn, W. (ed)** 1998. *The Experimenting Society: Essays in honor of Donald T. Campbell. Policy Studies Review Annual volume 11.* New Brunswick, New Jersey: Transaction Publishers.

Pielke Jr., R.A. **(**2007). *The Honest Broker: Making Sense of Science in Policy and Politics.* Cambridge, UK: Cambridge University Press.

Poteete, A.R., Janssen, M.A. & Ostrom, E. **(**2010). *Working together: collective actions, the commons, and multiple methods in practice*. New Jersey, USA: Princeton University Press.

Rodela, R. (2011). Social learning and natural resource management: the emergence of three research perspectives. *Ecology and Society, 16*(4), 30.

Rogers, P.J. & Williams, B. (2006). Evaluation for practice improvement and organizational learning. In Shaw, I.F., Greene, J.C. & Mark, M.M. (ed.), *Sage Handbook of Evaluation*. London, UK: Sage Publications Limited.

Rondinelli, D. **(**1993). *Development projects as policy experiments: an adaptive approach to development administration*. Routledge, London, Great Britain.

RWS (2009). *Harde werken met zachte trekken : voorbeelden van levende waterbouw*. http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2009/10/01/harde-werken-met-zachte-trekken-voorbeelden-van-levende-waterbouw.html

Sabatier, P. (1978). The acquisition and utilization of technical information by administrative agencies. *Administrative Science Quarterly, 23*, 286–411.

Sabatier, P. (1987). Knowledge, policy oriented learning and policy change: An advocacy coalition framework. *Science Communication, 8*(4), 649–692.

Sabel, C. & Zeitlin, J. (2011). Experimentalist governance. In Levi-Faur, D. (ed.), *The Oxford Handbook of Governance* (pp. 169-185). Oxford: Oxford University Press.

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. Public *Administration, 80*(1), 1–22.

Sanderson, I. (2009). Intelligent policy making for a complex world: pragmatism, evidence, and learning. *Political Studies, 57*(4), 699.

Schusler, T.M., Decker, D.J. & Pfeffer, M.J. (2003). Social learning for collaborative natural resource management. *Society and Natural Resources, 15*, 309–326.

Smith, A. (2007). Translating sustainabilities between green niches and socio-technical regimes. *Technology Analysis & Strategic Management, 19*(4), 427–450.

Star, S.L. & Griesemer, J.R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's museum of vertebrate zoology, 1907-39. *Social Studies of Science, 19*(3), 387–420.

Steffen, W. (2009). Interdisciplinary research for managing ecosystem services. *Proceedings for the National Academy of Sciences, 106*(5), 1301–1302.

Stoker, G. (2010). Exploring the promise of experimentation in political science: micro-foundational insights and policy relevance. *Political Studies, 58,* 300–319.

Termeer, C., Dewulf, A., van Rijswick, H., van Buuren, A., Huitema, D., Meijerink, S., Rayner, T. & Wiering, M. (2011). The regional governance of climate adaptation: A framework for developing legitimate, effective, and resilient governance arrangements. *Climate Law, 2*, 159–179.

Vedung, E. **(**1997). *Public policy and program evaluation.* New Brunswick, USA: Transaction Publishers.

Voß, J. & Bornemann, B. **(**2011). The politics of reflexive governance: challenges for designing adaptive management and transition management. *Ecology and Society, 16*(2), 9.

Vreugdenhil, H., Slinger, J., Thissen, W. & Ker Rault, P. (2010). Pilot projects in water management. *Ecology and Society, 15*(3), 13.

Walters, C.J. & Hollings, C.S. (1990). Large-scale management experiments and learning by doing. *Ecology, **71***, 2060–2068.

Weber, M. (1968). Economy and *society: An outline of interpretative sociology*. New York, USA: Bedminster Press.

Webler, T., Kastenholz, H. & Renn, O. (1995). Public participation in impact assessment: a social learning perspective. *Environmental Impact Assessment Review, 15*, 443–463.

Weiss, C. (1977). Research for policy's sake: the enlightenment function of social research. *Policy Analysis, 3*(4), 531–545.